Flipkart Mobile Phone Data Analysis

Introduction

This document provides a detailed analysis of a Flipkart dataset containing information about mobile phones, with a focus on models priced under ₹20,000. The primary goal of this analysis is to identify trends and provide recommendations based on key features like price, discount, ratings, battery life, and network type.

The analysis workflow includes:

- **Data Collection:** Gathering real-world data from the Flipkart website.
- **Data Cleaning and Preprocessing:** Handling missing values and converting data types.
- **Feature Engineering:** Creating new features to categorize data.
- **Exploratory Data Analysis (EDA):** Visualizing data to uncover key insights.
- **Recommendations:** Providing a summary of the findings.

Data Collection

The real-world data used for this analysis was collected by applying web scraping techniques to the Flipkart website. The process specifically targeted mobile phones available for under ₹20,000, retrieving information on various models, their prices, discounts, ratings, and other key specifications.

Data Cleaning and Preprocessing

The initial step involves loading the dataset and examining its structure. After loading the data, a check for null values revealed missing information in the **Rating**, **Network type**, **OSI**, and **Camera** columns.

- The **price** and **discount** columns were cleaned by removing special characters and converting them to numerical data types.
- The **Rating** column, which had a few missing values, was imputed with the mean of the available ratings to maintain data integrity.
- A new feature called **Status** was created to categorize phones based on battery and network type, with phones having a battery life greater than 1000 mAh on a 2G network or greater than 5000 mAh on a 5G network being labeled as "**Suggested**".
- The **OSI** (Operating System Interface) column, which had a significant number of missing values, was filled with "**Keypad Mobiles**" to distinguish between smartphones and feature phones. A new **Type** column was then created to explicitly categorize the phones as either "Smart Mobiles" or "Keypad Mobiles."
- The **Brand** column was extracted from the **Model** column for a clearer analysis of brand performance.

Exploratory Data Analysis (EDA)

The visualization phase provides a clear picture of the data distribution. The first plot shows the count of suggested versus not suggested phones.

Next, an analysis of the brand distribution among the suggested phones was performed, highlighting the brands that meet the specified criteria.

The final plot breaks down the affordable phones into "Smart Mobiles" and "Keypad Mobiles" by brand, providing a more granular view of the market. This helps in understanding which brands are catering to different user segments.

Summary of Findings

- The data cleaning process successfully transformed the raw data into a usable format by handling missing values and converting data types, which was crucial for accurate analysis.
- **POCO** emerged as a top performer among the "suggested" mobiles, which were defined by a good combination of battery and network type. This indicates that POCO is a strong contender for users who prioritize these features.
- In the "affordable" category, **REDMI** stands out as the leading brand. This is a user-friendly and budget-conscious brand, offering reliable quality at attractive prices, often with significant discounts.
- **ITEL** is identified as a strong brand in the **keypad phone** segment, providing essential features and dependable usability for customers who prefer simplicity.
- The analysis shows a significant number of **keypad phones** in the affordable category, particularly from brands like **Samsung** and **Kechaoda**.

This analysis provides a solid foundation for making informed decisions on which brands to recommend to customers, depending on their priorities—whether it's battery and network performance or a balance of affordability and features.